Sungani, H., Ngatunga, B., & Genner, M. J. (2016). Migratory behaviour shapes spatial genetic structure of cyprinid fishes within the Lake Malawi catchment. *Freshwater Biology*, *61*(7), 1062–1074. https://doi.org/10.1111/fwb.12767

Migratory behaviour shapes spatial genetic structure of cyprinid fishes within the Lake Malawi catchment

HAROLD SUNGANI[*,†], BENJAMIN P. NGATUNGA[‡] AND MARTIN J GENNER[*]

[*] School of Biological Sciences, Life Sciences Building, 24 Tyndall Avenue, Bristol. BS8 1TQ. United Kingdom.

[†]Fisheries Research Unit, PO Box 27, Monkey Bay, Malawi.

[‡]Tanzania Fisheries Research Institute, PO. Box 9750, Dar es Salaam, Tanzania.

Correspondence. Harold Sungani, School of Biological Sciences, Life Sciences Building, 24 Tyndall Avenue, Bristol. BS8 1TQ. United Kingdom. Email: harold.sungani@bristol.ac.uk

Key words. Microsatellite DNA, mitochondrial DNA, river fishes, conservation genetics, stock structure

SUMMARY

1. Genetic differences among freshwater fish populations are dependent on life history characteristics of the species, including the range of adult dispersal and the extent of homing to natal breeding grounds. However, the effects of variation in such characteristics on population genetic connectivity are rarely studied comparatively among closely related species.

2. We studied population genetic structure within three fish congeneric cyprinid species from the Lake Malawi catchment that differ substantially in life history traits and conservation status, using a combination of microsatellite and mitochondrial DNA markers. Mpasa (*Opsaridium microlepis*) is a large (70 cm total length) migratory species that spawns in rivers, but as an adult is exclusively known from the main lake body. Sanjika (*Opsaridium microcephalum*), is medium size (30 cm total length) species that exists in lake breeding, river-lake migratory and apparently landlocked populations. Dwarf sanjika (*Opsaridium tweddleorum*) is a small non-migratory species (15 cm total length) that persists in small tributaries surrounding the main lake and adjoining rivers.

3. The results revealed striking differences among the three species in spatial genetic structuring. The river-lake migratory mpasa showed only weak yet significant population genetic structure within the main Lake Malawi catchment, suggesting that there is no strong natal homing. The habitat generalist sanjika showed only weak spatial genetic differentiation at microsatellite loci within the Lake Malawi catchment, but moderate structure in mitochondrial DNA, potentially reflecting male biased dispersal. The river restricted dwarf sanjika showed strong genetic structure in both microsatellite and mitochondrial DNA, suggesting strictly limited dispersal at both adult and juvenile stages.

4. We conclude that contrasting migration life-histories have resulted in dramatically different patterns of population genetic structure among these congeneric species. The observed patterns demonstrate how divergent life history evolution may strongly influence broader patterns of population genetic connectivity in freshwater fish, with consequences for management and conservation. Specifically the results suggesting gene flow among Lake Malawi populations of mpasa, an IUCN red-listed "Endangered" species endemic to the lake catchment, implying that conservation initiatives operating at both local and catchment scales are needed to reverse local population decline.

Introduction

Conservation and management of freshwater fish species can benefit from an understanding of the genetic connectivity of populations which is dependent on multiple factors, including the rates of migration between populations and effective population sizes (Palumbi, 2003). Migration rates are fundamentally governed by movement behaviour throughout the life history of individuals. In many pelagic freshwater fish species an important dispersal phase is present during the egg and post-hatching larval stages. However, in most benthic freshwater or anadromous species dispersal is primarily restricted to adult phases. Thus, in freshwater riverine species that have benthic spawning habitats we may expect the degree of adult movement to be closely-linked to the extent of spatial population genetic structuring. However, it is also notable that in migratory species there may be cases where adults freely mix on lacustrine or marine feeding grounds, but return to natal riverine breeding grounds (Hedgecock *et al*., 2007). There are potentially strong benefits to breeding in rivers, due to a fundamentally lower predation risk to eggs and fry (Waldman *et al.,* 2008).

There is evidence that life history traits are closely correlated with the extent of population genetic connectivity. In multi-species analyses of marine fish data collected over large spatial scales, for example, the presence of a pelagic larval dispersal stage leads to lower population genetic structuring (Riginos *et al.,* 2011; Selkoe *et al.,* 2014). Over narrower taxonomic and spatial scales a potentially powerful method of inferring the role of life histories in determining patterns of population genetic structure is to directly compare closely related species with contrasting life history traits. In fishes, evolutionary switches between migratory and non-migratory behavioural patterns have occurred multiple times. Such switches have been particular common in salmonids and clupeids (e.g. alosines), which can have both migratory and resident populations of the same species (Fleming 1996, Roff 1988). In these species groups, it is common for the populations containing migratory individuals to exhibit reduced genetic structuring relative to the exclusively non-migratory populations (e.g. Hindar *et al.,* 1991; Jolly *et al.,* 2011).

Lake Malawi (= Lake Nyasa) in East Africa is an ancient lake estimated to have formed following rifting that initiated 8-12 million years ago (Mortimer *et al.,* 2007; Danley *et al.,* 2012). It has a maximum depth of 700 m, is 580 km long and 30–80 km wide. The lake has multiple inflowing rivers and a single outflow into the Zambezi drainage. A series of waterfalls along this outflow, the Shire River, form a biogeographic barrier between the Lake Malawi and Zambezi fish faunas (Tweddle & Willoughby 1979). We studied three species of river-spawning cyprinid from the Lake Malawi catchment that differ substantially in life history traits and conservation status. Mpasa (*Opsaridium*

microlepis) is a large sized (70cm maximum total length) migratory species that spawns in rivers but as an adult is otherwise exclusively found in the main lake body (Tweddle, 1987). The species is endemic to Lake Malawi catchment and is IUCN red-listed as "Endangered" due to apparent declines in abundance linked to fishing and alteration of spawning habitat (Kazembe *et al.*, 2006a). A second species, sanjika (*Opsaridium microcephalum*), is medium sized species (30cm maximum total length) that exists in lake breeding, river-lake migratory and apparently landlocked populations (Tweddle, 1983; Tweddle & Turner 2014). The species is also endemic to Lake Malawi and has a conservation status of "Vulnerable" (Kazembe *et al.*, 2006b). A third species, *Opsaridium tweddleorum*, is a small bodied species (15 cm maximum total length) that persists only in streams and small rivers (Skelton 2001). It is not endemic to Lake Malawi, as it is also found in tributaries of the Zambezi outside the Lake Malawi catchment. It has a conservation status of "Least Concern" (Vreven, 2006). Here we used a combination of microsatellite and mitochondrial DNA markers to quantify the extent of population connectivity in these species with contrasting migration behaviour, with a view to providing data to inform management and conservation of these species and critical habitats.

METHODS

Study sites, sample collection and DNA extraction.

Sixteen sites were sampled across the Lake Malawi catchment between 4th of May 2010 and 28th of January 2014 (Figure 1). Samples were obtained by fishing using seine nets, purchasing from artisanal fishermen, or from local markets (Tables 1 & 2). DNA samples were collected as either fin clips or whole juvenile fish, and were preserved in 100% ethanol. Total genomic DNA was extracted using the Wizard genomic DNA purification kit (Promega), following the protocol of the manufacturer.

Microsatellite DNA analyses

PCR reactions were performed in a multiplex of ten microsatellite loci (Sungani & Genner 2015; Table S1). Each PCR contained 5 µL of QIAGEN Multiplex mix, 1 µL of DNA, 0.5 µL of each primer (10 µM concentration), and 3 µL of RNase-free water. PCR conditions were as follows: 5 min at 95°C, then 35 cycles of 94°C for 1 min, 58°C for 90 s and 72°C for 1 min, followed by 72°C for 15 min. Amplification success was checked on 1% agarose gel. Amplified fragments were separated on an ABI3500 sequencer alongside a Genescan500 size standard. Alleles were identified and sizes checked using GeneMapper4.0 (Applied Biosystems Inc., Foster City, CA, USA).

Allele frequencies, observed (H_o) and expected (H_e) heterozygosity (Nei 1987) were calculated with Arlequin 3.5 (Excoffier & Lischer 2010). Deviations from Hardy-Weinberg equilibrium (HWE) were

examined for each population at each locus by using Exact tests with Markov chain permutations (chain length 1,000,000 with 100,000 dememorization steps) in Arlequin 3.5. Microchecker 2.2.3 (van Oosterhout *et al.*, 2004) was used to check for the presence of null alleles. Genetic differentiation between populations was estimated by calculating Wright's F_{ST} statistic (Weir & Cockerham 1984) and the significance of population genetic differences was tested using Exact with 1000 permutations in Genepop 4.2 (Rousset 2008). Tests of the power of the combined microsatellite loci for each species to detect significant spatial genetic differences were conducted using Powsim (Ryman & Palm, 2006). Specifically, we used the proportion of Fisher's tests that were significant ($P < 0.05$) following 500 runs, effective population sizes of 100 or 500, and parameters that provided estimated $F_{ST} = 0.005$ and $F_{ST} = 0.010$. We simulated tests of differences between two populations, each with 22 individuals which was the mean sample size of our study, and using the total allele frequencies observed across all populations. The extent of genetic differences among populations was visualised using multidimensional scaling in the MASS package (Ripley et al. 2015) in R 3.2.1 (R Development Core Team 2015). Population structure was analysed in each species independently using a model-based clustering algorithm with STRUCTURE 2.3 (Pritchard *et al.*, 2000). Ten runs were applied for each *K*. We applied a model that allowed for admixture with correlated allele frequencies, and conducted runs both with and without the Location Prior employed. Each run had a burn-in of 100,000 iterations followed by 100,000 MCMC replicates. STRUCTURE HARVESTER (Evanno *et al.*, 2005) and CLUMPAK (Kopelman *et al.*, 2015) were used to calculate summary data from the STRUCTURE runs.

Mitochondrial DNA analyses

The primers LCO1490 and HCO2198 (Folmer *et al.*, 1994) were used to amplify a ~700bp fragment of the *cytochrome oxidase subunit 1* (*COI*) gene. PCR was carried out in 25 µl volumes, including 12.5 µl of MyTaq (Bioline), 0.5 µl of each primer, 10.5 µl of H_2O, and 1 µl of DNA. PCR conditions were as follows: 4 min at 95 °C; then 34 cycles of 95 °C for 30 s, 50 °C for 30 s and 72 °C for 30s, followed by 72 °C for 7 min. Amplification success was checked on 1% agarose gel Sequencing of the PCR products was outsourced to Macrogen (Seoul, Korea). Sequences were aligned using ClustalW in DAMBE 5.3.32 (Xia, 2013). Haplotype networks were constructed using Hapstar (Teacher & Griffiths 2011), based on minimum spanning trees generated in Arlequin 3.5. Haplotype (h) and nucleotide (π) diversities were estimated per locality with DNASP 5.0 (Librado & Rozas, 2009). Genetic differentiation among pairs of populations was estimated using Φ_{ST}, and the significance of differences was tested using 1000 permutations in Arlequin 3.5, while multidimensional scaling as described above was used to visualise Φ_{ST} among populations. New sequences generated have the Genbank Accession numbers KU941588 to KU941831.

Results

Microsatellite data

Loci that were monomorphic within a species were excluded from further analyses, namely Ops20 in *O. microlepis*, and Ops 5 and Ops7 in *O. microcephalum* (Table S1). In *O. microlepis* across the 6 populations and 9 loci, we conducted 44 population-level tests of deviation from HWE (Table S1). Of these seven were significant ($P \leq 0.05$), five of which were at locus Ops16 and two were at locus Ops3. In *O. microcephalum* across 11 populations and 8 loci, we conducted 53 population-level tests of deviation from HWE (Table S1). Of these eight were significant ($P \leq 0.05$), six of which were at locus Ops16 and two at Ops3. In *O. tweddleorum* across 6 populations 10 loci, we conducted 42 population-level tests of deviation from HWE. Of these 11 were significant ($P \leq 0.05$), three of which were at locus Ops16 (Table S1). Significant deviations from HWE at Ops16 and Ops3 were in all cases associated with the potential presence of null alleles, as indicated by Microchecker (Table S2).

In *O. microlepis*, 202 tests of linkage disequilibrium between pairs of loci were performed, of which nine were significant ($P \leq 0.05$). In *O. microcephalum*, 106 tests of linkage disequilibrium between pairs of loci were performed, of which 1 was significant ($P \leq 0.05$). In *O. tweddleorum*, 116 tests of linkage disequilibrium between pairs of loci were performed, of which 4 were significant ($P \leq 0.05$). Given a lack of evidence for consistent patterns of linkage among loci across species all remaining loci were retained (Table S3). There were substantial differences among species in the power of exact tests on microsatellite data to detect significant differences among populations. At F_{ST} of 0.005 and 0.01, all species had low probability of rejecting the null hypothesis of panmixia (< 0.80). At F_{ST} of 0.05, O. *microlepis* and *O. tweddleorum* were able to reliably reject panmixia (probability > 0.99), while *O. microcephalum* performed less well (probability ~ 0.82)(Table S4).

A global test revealed an overall significant genetic difference among *O. microlepis* populations (F_{ST} = 0.063, Exact test, $P < 0.001$), and there were significant differences ($P \leq 0.05$) among all pairs of populations with the exception of populations from the North Rukuru and Songwe rivers, and the North Rukuru and Linthipe rivers (Table 3). The Lake Malombe population differed substantially from all main Lake Malawi populations (F_{ST} range = 0.172 to 0.217), while there was considerably less genetic structure present among the Lake Malawi populations (F_{ST} range 0.009 to 0.055; Fig 2a; Table 3). One locus (Ops11) showed strong F_{ST} among populations, while at the other eight loci F_{ST} was less than 0.01 (Figure S1). Structure resolved $K = 3$ as the most likely number of populations for analyses with

and without the location prior (Table S5). This analysis showed that the Lake Malombe population represented a genetically different population to the five from the main Lake Malawi catchment (Fig. 3a), but there was no clear pattern of structuring within the Lake Malawi catchment. There was no significant genetic isolation by distance (r = 0.34, P = 0.25).

A global test revealed an overall significant difference among *O. microcephalum* populations (F_{ST} = 0.012, Exact test $P < 0.001$), and there were 17 significant differences ($P \leq 0.05$) in the 55 pairwise tests (Table 4). The Kamuzu Dam and Lake Malombe populations were significantly different from each other and most populations in the Lake Malawi catchment. F_{ST} values ranged between -0.0215 and 0.183 (Fig. 2c, Table 4). Mean F_{ST} among populations was greatest at Ops11, Ops 16 and Ops19, but all loci showed mean $F_{ST} < 0.01$ (Fig. S1). Structure resolved K = 2 as the most likely number of populations for analyses with the location prior (Table S4), but there was no evidence of two discrete groups of individuals in the Structure output (Fig. 3b). Analyses without the location prior indicated K = 1 as the most likely number of populations (Table S5). There was no evidence of significant genetic isolation by distance (r = 0.005, P = 0.387).

A global test revealed an overall significant difference among *O. tweddleorum* populations (F_{ST} = 0.437, Exact test $P < 0.001$), and there were significant differences ($P \leq 0.05$) among all pairs of populations (Table 5). F_{ST} values ranged between 0.139 and 0.609 (Fig. 2e, Table 5). Mean F_{ST} was high (> 0.1) at all loci, with the exceptions of Ops3 and Ops20 (Fig. S1). Structure resolved K = 6 as the most likely number of populations for analyses with and without the location prior (Fig. 3c; Table S5). Under both sets of analyses all populations were largely distinct, with the exception of the neighbouring Lingadzi and Lumbadzi populations. There was no significant genetic isolation by distance (r = 0.23, P = 0.23).

Mitochondrial DNA data

The length of the *COI* sequences screened were 541 bp for *O. microlepis*, 579 bp for *O. microcephalum* and 574 bp for *O. tweddleorum*. In *O. microlepis* only two haplotypes differing by 1 bp were identified from the 98 individuals sequenced (Fig. 4a). There were no overall significant genetic difference among *O. microlepis* populations (Φ_{ST} = 0.002, P = 0.320). Pairwise comparisons of populations revealed Φ_{ST} values ranging between -0.081 and 0.157, and only one significant difference was present ($P < 0.05$), between the Bua and North Rukuru populations (Fig. 2b, Table 3). In *O. microcephalum* a total of 26 haplotypes were found among the 88 individuals sequenced, with several populations exhibiting private haplotypes (Fig. 4b). Overall population genetic differentiation was highly significant in *O. microcephalum* (Φ_{ST} = 0.088, $P < 0.001$). Pairwise comparisons of populations revealed Φ_{ST} values

ranging between -0.021 and 0.259, and six of the ten pairwise comparisons were significant ($P < 0.05$), all involving either the Linthipe or Chia rivers (Fig. 2d, Table 4). In *O. tweddleorum* five haplotypes were identified among the 57 individuals sequenced (Fig. 3), with only one haplotype shared between multiple populations, all of which were geographically located in the south-east of the Lake Malawi catchment. In a global analysis there was a highly significant difference identified among *O. tweddleorum* populations ($\Phi_{ST} = 0.416$, $P < 0.001$). Ten of the fifteen pairwise comparisons were significant ($P < 0.05$), with Φ_{ST} values ranging between 0 and 1 (Fig 2f, Table 5). There were no significant associations between genetic distance (Φ_{ST}) and geographic distance for any of the species (Mantel tests: *O. microlepis*, $r = 0.645$, $P = 0.148$; *O. microcephalum* $r = 0.386$, $P = 0.253$; *O. tweddleorum* $r = 0.238$, $P = 0.123$).

Discussion

The results revealed striking differences among the three focal species in spatial genetic structuring. The large bodied river-lake migratory mpasa (*O. microlepis*) showed weak yet significant genetic structuring among populations from the main Lake Malawi catchment in microsatellite DNA, but no evidence of structuring in mitochondrial DNA. The biological relevance of low yet statistically significant genetic structuring is an open to debate, but such patterns are consistent with partial reproductive isolation (Knutsen et al. 2010). At the very least the results are suggestive of a lack of strong natal homing in mpasa. Nevertheless there may be some dispersal limitation among populations, perhaps related to adult mpasa remaining in the vicinity of natal sites as adults.

Notably, we did find some evidence of relatively strong spatial population genetic structure between the Lake Malawi and Lake Malombe mpasa populations sampled. This is indicative of a lack of dispersal through the connecting Upper Shire River (Fig. 1), however we note that our sample size from the Lake Malombe population was relatively small ($n = 14$). Moreover, all individuals sampled were juveniles collected from the same location on the same day, and the population contained low heterozygosity relative to the Lake Malawi populations. Thus, we cannot rule out the possibility that the observed structure is a result of sampling only very closely related individuals. Notably mpasa is very rare in the vicinity of Lake Malombe, although juveniles have previously been observed in a tributary of Shire River at Liwonde (D. Tweddle, pers. comm). It is possible that there is a breeding population within Lake Malombe, but further work is required to determine if this is genetically distinct from those in Lake Malawi.

The weak population genetic structure among Lake Malawi mpasa populations in our results contrasts with results of Changadeya *et al.,* (2013) who also used microsatellite loci to study population genetic differences among mpasa populations from the Linthipe, Bua, Dwangwa and North Rukuru systems. Although qualitatively similar in demonstrating significant genetic differences among populations, Changadeya *et al.,* (2013) reported consistently high F_{ST} values among all populations (mean 0.17, range 0.14 to 0.19). They also reported highly significant deviations from HWE in all 20 loci they studied, and in all populations. They observed excess heterozygosity relative to expected values, which the authors proposed may be due to outbreeding. Our data suggest are not consistent with this interpretation, and instead suggest that populations of mpasa largely conform to HWE expectations.

We found only very weak population genetic structure within sanjika (*O. microcephalum*) using microsatellite makers, but some evidence of structure in mitochondrial DNA. The absence of genetic structure among many populations using microsatellite markers is likely to be linked to a low power of resolution (Table S4), driven by a low allelic diversity at most loci, when compared with mpasa and dwarf sanjika (Table S1). However, it is also potentially linked to broad adult dispersal and an opportunistic breeding strategy, as sanjika are known to spawn in both river and lacustrine environments (Tweddle & Turner, 2014). In contrast to the microsatellite diversity, mitochondrial DNA diversity was comparatively high in sanjika relative to other *Opsaridium* species studied. Significant population structure was present, but there was no evidence of a pattern of genetic isolation by distance. This result is partially compatible with male-biased dispersal in this species. Male-biased dispersal has been widely reported in fish using both genetic (Prugnolle & de Meeus, 2002) and tagging methods (Croft *et al.,* 2003), including migratory salmonids (Hutchings & Gerber, 2002). It is generally thought that the strategy has evolved to weaken male-male competition among siblings, and reduce probability of mating with closely-related females. Our data suggest there is scope to further investigate sex biases in the migratory behaviour of this species using molecular genetic markers, or perhaps trace-element analyses of otoliths (e.g. Walther & Thorrold, 2008).

The sampled populations of dwarf sanjika (*O. tweddleorum*) typically exhibited very strong evidence of genetic differentiation, and there was a significant pattern of genetic isolation by distance in the microsatellite data. These results strongly suggest that gene-flow among populations is infrequent and that the main lake is only very rarely used as a conduit for dispersal. It is interesting to speculate on reasons why dwarf sanjika does not possess an adult lacustrine dispersal phase equivalent to mpasa and sanjika. It is possible that the small body size enables year-round exploitation of smaller riverine food resources. This in turn may enable larger populations to persist in rivers than congeneric sympatric species meaning the costs of inbreeding are reduced.

The apparent contrast in mtDNA and microsatellite diversity in the large bodied *Opsaridium* species is intriguing. Mpasa had low mtDNA diversity, yet relatively high microsatellite diversity. Sanjika had comparatively higher mtDNA diversity, yet low microsatellite diversity. Such patterns may be reflective of chance events during population bottlenecks during megadroughts that have repeatedly desiccated Lake Malawi over the last 1.3 million years (Lyons *et al.*, 2015). However further investigations using many more markers are required to determine if the patterns are consistent across the genome.

Connectivity and conservation

Integrating information on evolutionary connectivity between populations in the design of a network of designated areas can, in principle, increase the efficiency of biodiversity protection. Management priorities can be based on their degree of diversity and evolutionary uniqueness with respect to other populations (Woodruff, 1989; Crozier & Kusmierski, 1994). Specifically, in terms of fisheries management, information on the extent of genetic connectivity can help to reveal aspects of the underlying dispersal biology of the species and help to identify management units (Palsbøll *et al.*, 2007). Our data strongly imply that three *Opsaridium* species have different genetic structures governed by their life histories and ecological preferences. Management therefore needs to recognise these issues, in addition to threats from fisheries, habitat fragmentation, pollution and water abstraction (Vörösmarty *et al.*, 2010).

The IUCN 'Endangered' mpasa and 'Vulnerable' sanjika have traditionally been exploited during the annual spawning runs, and both species continue to be the focus of fisheries in rivers in both Malawi and Tanzania. Although few quantitative data are available on catches, there is a collective opinion that catches of both have declined (Kazumbe *et al.*, 2006a, 2006b). Mpasa, in particular are suggested to have declined substantially in all major rivers studied between the 1940s and 1970s, based on interviews with fishers (Tweddle, 1981). Specific threats include overfishing, degradation of spawning habitat, weir construction and fish poisoning (Tweddle, 2001). An important challenge for the long-term conservation of both species is to protect the remaining intact river habitat. Sampling for this study has demonstrated that number of rivers including the Bua, Linthipe North Rukuru and Songwe still have spawning runs with intact breeding habitat for adult fish of both species. Our results are suggestive of partially restricted gene flow among populations in both species and imply that conservation initiatives that protect spawning and juvenile habitat in individual rivers may have both local and broader geographic benefits for rebuilding of the total Lake Malawi spawning stocks of these species. Such habitat protection should be coupled with regular monitoring of abundance of these species within

rivers. Since juvenile *Opsaridium* are sensitive to changes in water quality, food resources, river flow, turbidity and temperature (Tweddle, 1983), flourishing populations of juveniles could be useful indicators of river ecosystem health.

Finally, our results are consistent with field observations that the dwarf sanjika is river-restricted and non-migratory. It is therefore representative of many freshwater fish species of small rivers and ephemeral habitats across Africa, and notably the high degree of genetic structure in *O. tweddleorum* is mirrored over similar spatial scales in many other fish groups including cyprinids (*Pseudobarbus* sp.; Swartz *et al.,* 2009), catfishes (*Chiloglanis* sp., Schmidt *et al.,* 2014), killifishes (*Nothobranchius furzeri*; Bartáková *et al.,* 2013) and cichlids (*Astatotilapia calliptera*, Nichols *et al.,* 2015; *Pseudocrenilabrus* sp., Katongo *et al*. 2008; Egger et al. 2015). Together these results demonstrate a general pattern of only occasional dispersal and low genetic connectivity among many river systems in Africa, suggesting that local disturbance events or habitat alteration could strongly impact on the ecology of individual water bodies and lead to a loss of unique genetic diversity.

Acknowledgements

The work was supported by Commonwealth Scholarship MWCS-2012-249 to HS and a Royal Society-Leverhulme Trust Africa Award AA100023 and AA130107 to MG and BPN. Alan Smith, George Turner and Jennifer Swanstrom provided assistance with sample collection in Tanzania. Jane Coghill provided valuable support with genotyping the samples, while Alan Hudson is thanked for advice on Structure analyses.

References

Bartáková, V., Reichard, M., Janko, K., Polačik, M., Blažek, R., Reichwald, K., Cellerino, A. & Bryja, J. (2013) Strong population genetic structuring in an annual fish, *Nothobranchius furzeri*, suggests multiple savannah refugia in southern Mozambique. *BMC Evolutionary Biology*, **13**, 196.

Changadeya, W., Stima, H., Kassam, D., Kaunda, E. & Jere, W. (2013) Genetic diversity and structure of potamodromous *Opsaridium microlepis* (Günther) populations in the inlet rivers of Lake Malawi. *African Journal of Biotechnology* **12**, 6709-6717.

Croft, D. P., Albanese, B., Arrowsmith, B. J., Botham, M., Webster, M. & Krause, J. (2003) Sex-biased movement in the guppy (*Poecilia reticulata*). *Oecologia*, **137**, 62-68.

Crozier, R.H. & Kusmierski, R.M. (1994) Genetic distances and the setting of conservation priorities. In *Conservation Genetics* (eds Loeschcke, V., Tomiuk, J. & Jain, S. K), pp. 227–237. Birkhauser Verlag. Basel.

Danley, P. D., Husemann, M., Ding, B., DiPietro, L. M., Beverly, E. J. & Peppe, D. J. (2012) The impact of the geologic history and paleoclimate on the diversification of East African cichlids. *International Journal of Evolutionary Biology*, **2012**. Article ID 574851.

Egger, B., Klaefiger, Y., Indermaur, A., Koblmüller, S., Theis, A., Egger, S., Näf, T., Van Steenberge, M., Sturmbauer, C., Katongo, C. & Salzburger, W. (2014) Phylogeographic and phenotypic assessment of a basal haplochromine cichlid fish from Lake Chila, Zambia. *Hydrobiologia*, **748**, 171-184.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier, L. & Lischer, H.E. (2010) Arlequin Suite Ver. 3.5: A new series of programs to perform population genetic analyses under Linux and Windows. *Molecular Ecology Resources* 10, 564-567.

Fleming, I.A. (1996). Reproductive strategies of Atlantic salmon: ecology and evolution. *Reviews in Fish Biology and Fisheries*, **6**, 379-416.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 1994, 294-299.

Hedgecock, D., Barber, P.H. & Edmands, S. (2007) Genetic approaches to measuring connectivity. *Oceanography* **20**, 70-79.

Hindar, K., Jonsson, B., Ryman, N. & Ståhl, G. (1991) Genetic relationships among landlocked, resident, and anadromous brown trout, *Salmo trutta* L. *Heredity*, **66**, 83-91.

Hutchings, J.A. & Gerber, L. (2002) Sex–biased dispersal in a salmonid fish. *Proceedings of the Royal Society of London B*, **269**, 2487-2493.

Jolly, M.T., Aprahamian, M.W., Hawkins, S.J., Henderson, P.A., Hillman, R., O'Maoiléidigh, N., Maitland P., Piper R. & Genner, M.J. (2012) Population genetic structure of protected allis shad (*Alosa alosa*) and twaite shad (*Alosa fallax*). *Marine Biology*, **159**, 675-687.

Katongo, C., Koblmüller, S., Duftner, N., Makasa, L. & Sturmbauer, C. (2005) Phylogeography and speciation in the *Pseudocrenilabrus philander* species complex in Zambian Rivers. *Hydrobiologia*, **542**, 221-233.

Kazembe, J., Magombo, Z., Khawa, D. & Kaunda, E. (2006) *Opsaridium microlepis*. The IUCN Red List of Threatened Species. Version 2015.1. <www.iucnredlist.org>. Downloaded on 18 June 2015.

Kazembe, J., Magombo, Z., Khawa, D. & Kaunda, E. (2006) *Opsaridium microcephalum*. The IUCN Red List of Threatened Species. Version 2015.1. <www.iucnredlist.org>. Downloaded on 18 June 2015.

Knutsen, H., Olsen, E.M., Jorde, P.E., Espeland, S.H., André, C. and Stenseth, N.C. (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular Ecology*, **20**, 768-783.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lyons, R.P., Scholz, C.A., Cohen, A.S., King, J.W., Brown, E.T., Ivory, S.J., Johnson, T.C., Deino, A.L., Reinthal, P.N., McGlue, M.M. and Blome, M.W. (2015) Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *Proceedings of the National Academy of Sciences* of the *USA*, doi: 10.1073/pnas.1512864112

Mortimer, E., Paton, D. A., Scholz, C. A., Strecker, M. R. & Blisniuk, P. (2007) Orthogonal to oblique rifting: effect of rift basin orientation in the evolution of the North basin, Malawi Rift, East Africa. *Basin Research*, **19**, 393-407.

Nichols, P., Genner, M.J., van Oosterhout, C., Smith, A., Parsons, P., Sungani, H. & Joyce, D.A. (2015). Secondary contact seeds phenotypic novelty in cichlid fishes. *Proceedings of the Royal Society of London B*, **282**, 20142272.

Palsbøll, P.J., Bérubé, M. & Allendorf, F.W. (2007) Identification of management units using population genetic data. *Trends in Ecology and Evolution* **22**, 11-16.

Palumbi, S.R. (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13**, 146-158.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Prugnolle F & de Meeus T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, **88**, 161-165.

Riginos, C., Douglas, K.E., Jin, Y., Shanahan, D.F. & Treml, E.A. (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, **34**, 566-575.

Ripley, B., Bates, D., Hornik, K., Gebhardt, A. & Firth, D. (2015) Package 'MASS'. Available at: cran.r-project.org/web/packages/MASS (accessed 18 December 2015).

R Development Core Team. (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Roff, D.A. (1988) The evolution of migration and some life history parameters in marine fishes. *Environmental Biology of Fishes*, **22**, 133-146.

Rousset, F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.

Ryman, N. & Palm, S. (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600–602.

Schmidt, R.C., Bart, H.L., Nyingi, D.W. & Gichuki, N.N. (2014) Phylogeny of suckermouth catfishes (Mochokidae: *Chiloglanis*) from Kenya: The utility of growth hormone introns in species level phylogenies. *Molecular Phylogenetics and Evolution*, **79**, 415-421.

Selkoe, K.A., Gaggiotti, O.E., Bowen, B.W. & Toonen, R.J. (2014) Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology*, **23**, 3064-3079.

Skelton, P.H. (2001) *A complete guide to the freshwater fishes of Southern Africa*. Struik. Cape Town.

Sungani, H. & Genner, M.J. (2015) Microsatellite markers from 454-sequencing of the threatened Lake Malawi endemic mpasa *Opsaridium microlepis*. In: *Microsatellite records for volume 7, issue 4*. *Conservation Genetic Resources* 7, 917-944.

Swartz, E.R., Skelton, P.H. & Bloomer, P. (2009) Phylogeny and biogeography of the genus *Pseudobarbus* (Cyprinidae): shedding light on the drainage history of rivers associated with the Cape Floristic Region. *Molecular Phylogenetics and Evolution*, **51**, 75-84.

Teacher, A.G.F. & Griffiths, D.J. (2011) HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources*, **11**, 151–153.

Tweddle, D. (1981) The importance of long-term data collection on river fisheries, with particular reference to the cyprinid *Opsaridium microlepis* (Günther, 1864) fisheries of the affluent rivers of Lake Malawi. In: *CIFA Technical Paper 8* (Ed. J.M. Kapetsky, J.M). pp. 145-163. FAO, Rome.

Tweddle, D. (1983) The fish and fisheries of Lake Malawi. *Journal of Science and Technology (Malawi)*, **4**, 55-83.

Tweddle, D. (1987) An assessment of the growth rate of mpasa, *Opsaridium microlepis* (Gunther, 1864) (Pisces: Cyprinidae) by length frequency analysis. *Journal of the Limnological Society of Southern Africa*, **13**, 52–57.

Tweddle, D. (2001) Threatened fishes of the world: *Opsaridium microlepis* (Günther, 1864) (Cyprinidae). *Environmental Biology of Fishes*, **61**, 72.

Tweddle, D. & Turner, G.F. (2014) Evidence for lacustrine breeding by sanjika *Opsaridium microcephalum* (Teleostei: Cyprinidae) in Lake Malawi. *African Journal of Aquatic Science*, **39**, 479-480.

Tweddle, D. & Willoughby N.G. (1979) The nature of the barrier separating the Lake Malawi and Zambesi fish faunas. An annotated checklist of the fish fauna of the river Shire south of

463 Kapachira Falls, Malawi. *Ichthyological Bulletin of the J.L.B. Smith Institute of Ichthyology*, **39**,
464 1-9.
465 Van Oosterhout, C., Hutchinson, W.F., Wills, D.P. & Shipley, P. (2004) MICRO-CHECKER: software
466 for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*,
467 **4**, 535-538.
468 Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden,
469 S., Bunn, S.E., Sullivan, C.A., Liermann, C.R. & Davies, P.M. (2010) Global threats to human
470 water security and river biodiversity. *Nature*, **467**, 555-561.
471 Vreven, E. (2006) *Opsaridium tweddleorum*. The IUCN Red List of Threatened Species. Version
472 2015.1. <www.iucnredlist.org>. Downloaded on 18 June 2015.
473 Waldman, J., Grunwald, C., & Wirgin, I. (2008). Sea lamprey *Petromyzon marinus*: an exception to
474 the rule of homing in anadromous fishes. *Biology Letters*, **4**, 659-662.
475 Walther, B.D., Thorrold, S.R., & Olney, J.E. (2008). Geochemical signatures in otoliths record natal
476 origins of American shad. *Transactions of the American Fisheries Society*, **137**, 57-69.
477 Weir, B., Cockerham, C. (1984) Estimating *F* statistics for the analysis of population structure.
478 *Evolution*, **38**, 1358–1370.
479 Woodruff, D.S. (1989) The problems of conserving genes and species. In: *Conservation for the twenty-*
480 *first century*. Eds. D. Western, and M. Pearl), pp. 76-88. Oxford University Press, New York.
481 Xia, X. (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology
482 and evolution. *Molecular Biology and Evolution*, **30**, 1720-1728.
483
484

485 **Figure Legends.**

486

487 **Fig. 1**. Sampling locations of the populations of three species *O. microlepis*, *O. microcephalum* and *O.*
488 *tweddleorum* within the Lake Malawi catchment. Colour shading indicates the species sampled from
489 each location (Table 1).

490

491 **Fig. 2**. Multidimensional scaling plots of genetic differences between populations, based on
492 microsatellite FST and mitochondrial Φ_{ST} values, for a) *O. microlepis*, b) *O. microcephalum* and c) *O.*
493 *tweddleorum*. of *O. microcephalum* data without a location prior indicated a single panmictic
494 population ($K = 1$).

495

496 **Fig. 3.** STRUCTURE assignment of individuals to populations using microsatellite data for a) *O.*
497 *microlepis*, b) *O. microcephalum* and c) *O. tweddleorum*. Analyses

498

499 **Fig. 4.** MtDNA (*COI*) haplotype network of sampled populations for a) *O. microlepis* b) *O.*
500 *microcephalum*, and c) *O. tweddleorum*.

Table 1: Sampled locations and sample size of populations of *O. microlepis*, *O. microcephalum* and *O. tweddleorum* analysed for nuclear (microsatellite) DNA

Species	Sampling Date	Location name	Sampling method	Latitude (S)	Longitude (E)	*N* individuals microsatellites
O. microlepis	28-01-2014	Songwe*	Seine nets	09º35.24'	033º46.16'	15
	28-01-2014	North Rukuru	Market	09º55.03'	033º55.68'	48
	25-01-2014	Dwangwa*	Seine nets	12º30.81'	034º06.97'	19
	24-01-2014	Bua	Artisanal fishers	12º47.19'	034º11.76'	19
	22-01-2014	Linthipe	Artisanal fishers	13º47.33'	034º26.09'	38
	04-05-2010	Malombe*	Seine nets	14º32.58'	035º12.66'	14
O. microcephalum	28-01-2014	Songwe*	Seine nets	09º35.24'	033º46.16'	5
	28-01-2014	North Rukuru	Seine nets	09º55.03'	033º55.68'	34
	26-01-2014	Ngala	Market	12º21.30'	034º03.57'	8
	25-01-2014	Dwangwa*	Seine nets	12º30.81'	034º06.97'	6
	24-01-2014	Bua	Artisanal fishers	12º47.19'	034º11.76'	56
	23-01-2014	Chia	Artisanal fishers	13º07.55'	034º19.64'	18
	23-08-2013	Lumbadzi*	Seine nets	13º47.64'	033º59.30'	18
	23-08-2013	Kamuzu Dam*	Artisanal fishers	14º10.47'	033º38.57'	20
	22-01-2014	Linthipe	Artisanal fishers	13º47.33'	034º26.09'	74
	20-01-2014	Monkey Bay	Market	14º04.17'	034º55.03'	16
	04-05-2010	Malombe*	Seine nets	14º32.58'	035º12.66'	6
O. tweddleorum	27-01-2014	North Rumphi	Seine nets	10º41.109'	034º10.883'	16
	09-09-2012	Ruhuhu	Seine nets	10º35.650'	034º39.20'	18
	26-01-2014	Khuyu	Seine nets	12º16.735'	033º59.939'	26
	23-01-2014	Lingadzi	Seine nets	13º32.061'	034º17.064'	10
	23-08-2013	Lumbadzi	Seine nets	13º47.64'	033º59.30'	8
	12-01-2011	Mtakataka	Seine nets	14º12.855'	034º30.819'	14

*Sample of juveniles, all other individuals of *O. microlepis* and *O. microcephalum* were adult size.

Table 2: MtDNA (*COI*) variation within populations of *O. microlepis*, *O. microcephalum* and *O.tweddleorum*.

Species	Population	*N* individuals mtDNA	Polymorphic sites	Number haplotypes	Haplotype diversity	Nucleotide diversity
O. microlepis	Linthipe	20	1	2	0.337	0.00062
	Bua	16	1	2	0.125	0.00023
	North Rukuru	47	1	2	0.459	0.00085
	Dwangwa	9	1	2	0.222	0.00041
	Malombe	6	0	1	0	0
	All 5 populations	98	1	2	0.352	0.00066
O. microcephalum	Linthipe	22	19	13	0.905	0.00613
	Chia	17	8	3	0.581	0.00593
	Bua	20	7	4	0.284	0.00198
	Ngala	7	8	4	0.714	0.00592
	North Rukuru	22	14	8	0.732	0.00544
	All 5 populations	88	31	25	0.726	0.00681
O. tweddleorum	Lumbadzi	8	0	1	0	0
	North Rumphi	16	1	2	0.462	0.00080
	Khuyu	26	0	1	0	0
	Lingadzi	10	0	1	0	0
	Mtakataka	14	0	1	0	0
	Ruhuhu	18	0	1	0	0
	All 6 populations	92	5	5	0.771	0.00314

Table 3: Genetic differences among *Opsaridium microlepis*. Below the diagonal are F_{ST} and Φ_{ST} values for microsatellite and mitochondrial data, respectively. Above the diagonal *P*-values

Microsatellite	Bua	Dwangwa	Linthipe	North Rukuru	Songwe	Malombe
Bua	-	< 0.001	0.006	< 0.001	0.006	< 0.001
Dwangwa	0.025	-	0.002	< 0.001	0.002	< 0.001
Linthipe	0.002	0.027	-	0.038	0.004	< 0.001
North Rukuru	0.018	0.055	0.009	-	0.371	< 0.001
Songwe	0.024	0.036	0.014	0.002	-	< 0.001
Malombe	0.172	0.217	0.200	0.211	0.198	-

MtDNA	Bua	Dwangwa	Linthipe	North Rukuru	Malombe
Bua	-	0.455	0.630	0.040	0.999
Dwangwa	0.023	-	0.940	0.075	0.999
Linthipe	-0.012	-0.048	-	0.110	0.651
North Rukuru	0.140	0.105	0.040	-	0.187
Malombe	-0.079	-0.051	-0.081	0.157	-

Table 4: Genetic differences among *Opsaridium microcephalum*. Below the diagonal are F_{ST} and Φ_{ST} values for microsatellite and mitochondrial data, respectively. Above the diagonal *P*-values.

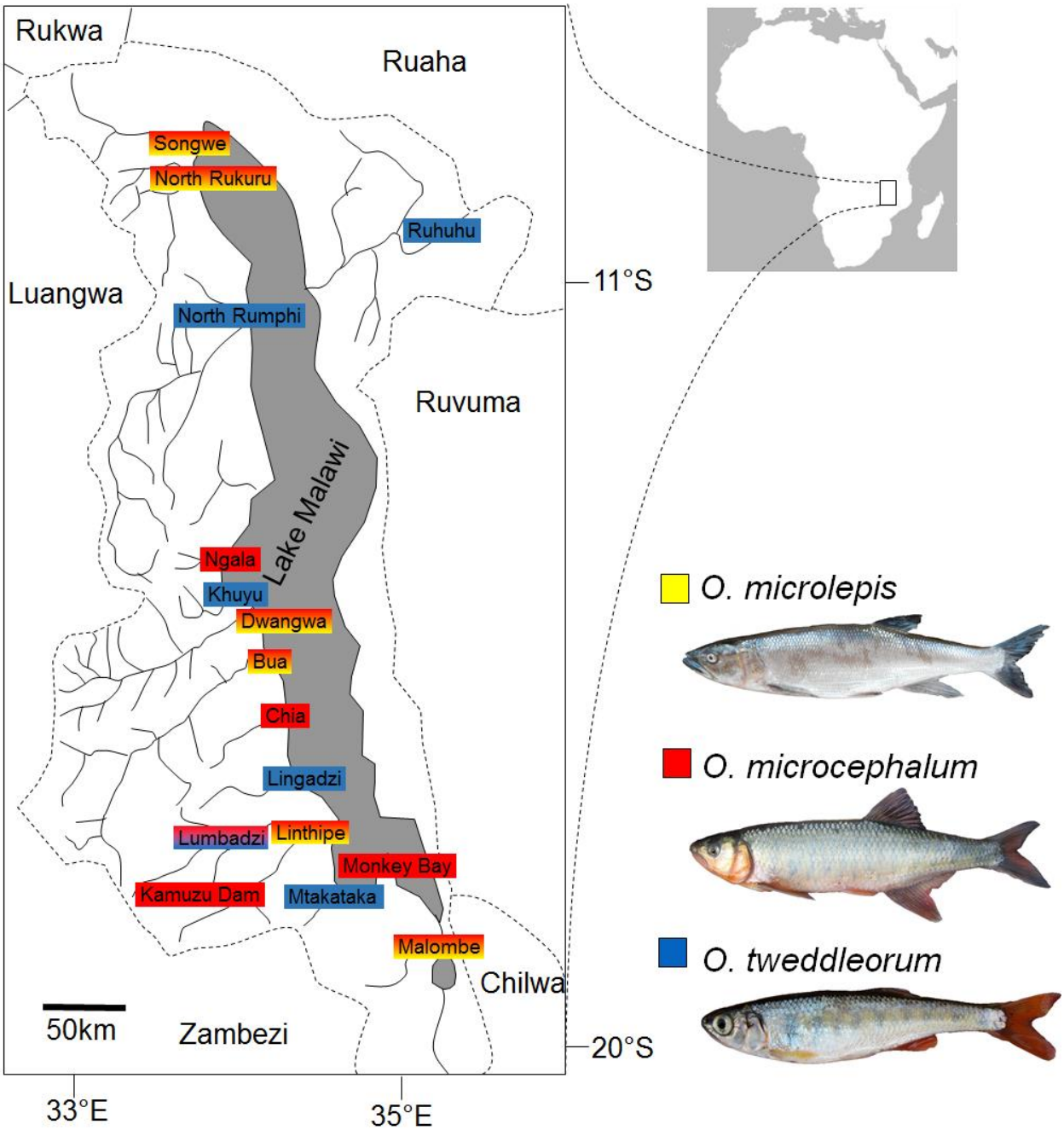
Microsatellite	Bua	Chia	Dwangwa	Kamuzu Dam	Linthipe	Lumbadzi	Monkey Bay	Ngala	North Rukuru	Songwe	Malombe
Bua	-	0.139	0.780	0.033	0.585	0.140	0.112	0.668	0.915	0.468	0.015
Chia	0.005	-	0.759	0.025	0.166	0.193	0.250	0.496	0.929	0.318	0.016
Dwangwa	-0.005	-0.017	-	0.006	0.974	0.875	0.796	0.287	0.893	0.985	0.034
Kamuzu Dam	0.049	0.039	0.094	-	0.077	< 0.001	0.139	0.004	0.038	0.039	< 0.001
Linthipe	-0.003	0.005	-0.014	0.048	-	0.032	0.238	0.265	0.490	0.752	0.004
Lumbadzi	0.006	0.015	-0.019	0.087	0.010	-	0.291	0.013	0.288	0.514	0.004
Monkey Bay	0.013	0.001	-0.001	0.017	0.009	0.019	-	0.087	0.190	0.618	0.003
Ngala	0.029	-0.011	0.049	0.035	0.029	0.067	0.026	-	0.771	0.225	0.257
North Rukuru	-0.005	-0.010	-0.007	0.046	0.002	0.013	0.009	0.001	-	0.576	0.042
Songwe	-0.014	-0.008	-0.031	0.057	-0.022	-0.009	-0.009	0.018	-0.020	-	0.428
Malombe	0.081	0.090	0.107	0.183	0.079	0.078	0.134	0.041	0.069	0.050	-

MtDNA	Bua	Chia	Linthipe	Ngala	North Rukuru
Bua	-	< 0.001	< 0.001	0.587	0.277
Chia	0.066	-	< 0.001	0.256	0.001
Linthipe	0.181	0.118	-	0.022	< 0.001
Ngala	-0.021	0.011	0.144	-	0.172
North Rukuru	0.004	0.081	0.259	0.041	-

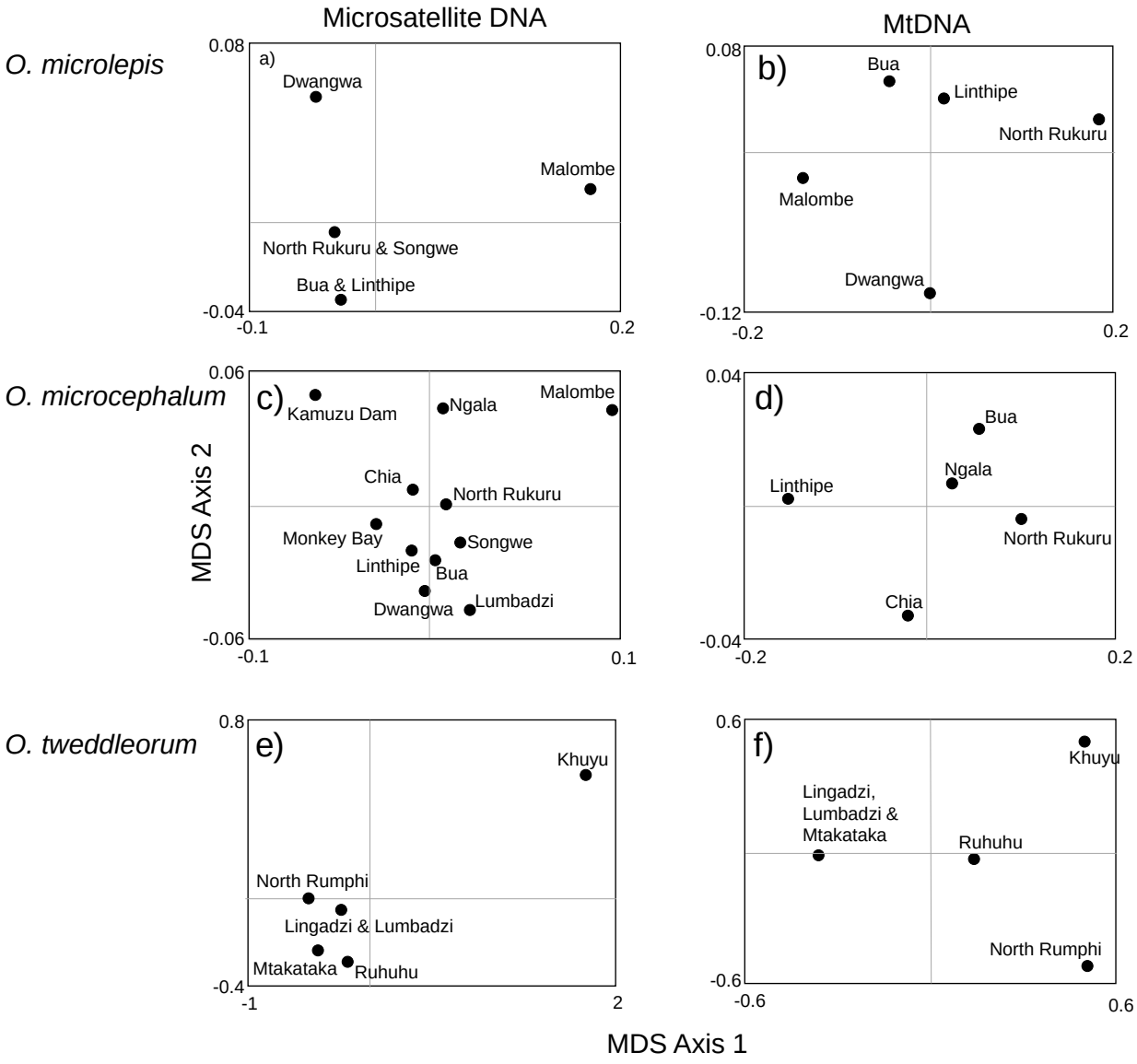
Table 5: Genetic differences among *Opsaridium tweddleorum*. Below the diagonal are F_{ST} and Φ_{ST} values for microsatellite and mitochondrial data, respectively. Above the diagonal *P*-values

Microsatellite	Khuyu	Lingadzi	Lumbadzi	North Rumphi	Mtakataka	Ruhuhu
Khuyu	-	0.005	0.033	< 0.001	< 0.001	0.085
Lingadzi	0.485	-	0.190	0.001	< 0.001	0.002
Lumbadzi	0.453	0.139	-	0.014	0.011	< 0.001
North Rumphi	0.609	0.362	0.274	-	0.007	< 0.001
Mtakataka	0.562	0.326	0.267	0.274	-	< 0.001
Ruhuhu	0.528	0.396	0.318	0.445	0.390	-

MtDNA	Khuyu	Lingadzi	Lumbadzi	North Rumphi	Mtakataka	Ruhuhu
Khuyu	-	< 0.001	0.009	< 0.001	< 0.001	< 0.001
Lingadzi	1	-	0.991	< 0.001	0.991	< 0.001
Lumbadzi	1	0	-	< 0.001	0.991	0.234
North Rumphi	1	1	1	-	< 0.001	< 0.001
Mtakataka	1	0	0	1	-	0.315
Ruhuhu	0.313	0.122	0.084	0.196	0.083	-



528
529
530 Fig. 1


Microsatellite DNA
MtDNA
O. microlepis
O. microcephalum
O. tweddleorum
a)
Dwangwa
Malombe
North Rukuru & Songwe
Bua & Linthipe
0.08
-0.04
-0.1
0.2
b)
Bua
Linthipe
North Rukuru
Malombe
Dwangwa
0.08
-0.12
-0.2
0.2
c)
Kamuzu Dam
Ngala
Malombe
Chia
North Rukuru
Monkey Bay
Songwe
Linthipe
Bua
Dwangwa
Lumbadzi
MDS Axis 2
0.06
-0.06
-0.1
0.1
d)
Bua
Ngala
Linthipe
North Rukuru
Chia
0.04
-0.04
-0.2
0.2
e)
Khuyu
North Rumphi
Lingadzi & Lumbadzi
Mtakataka
Ruhuhu
0.8
-0.4
-1
2
f)
Khuyu
Lingadzi,
Lumbadzi &
Mtakataka
Ruhuhu
North Rumphi
0.6
-0.6
-0.6
0.6
MDS Axis 1

531
532 Fig. 2

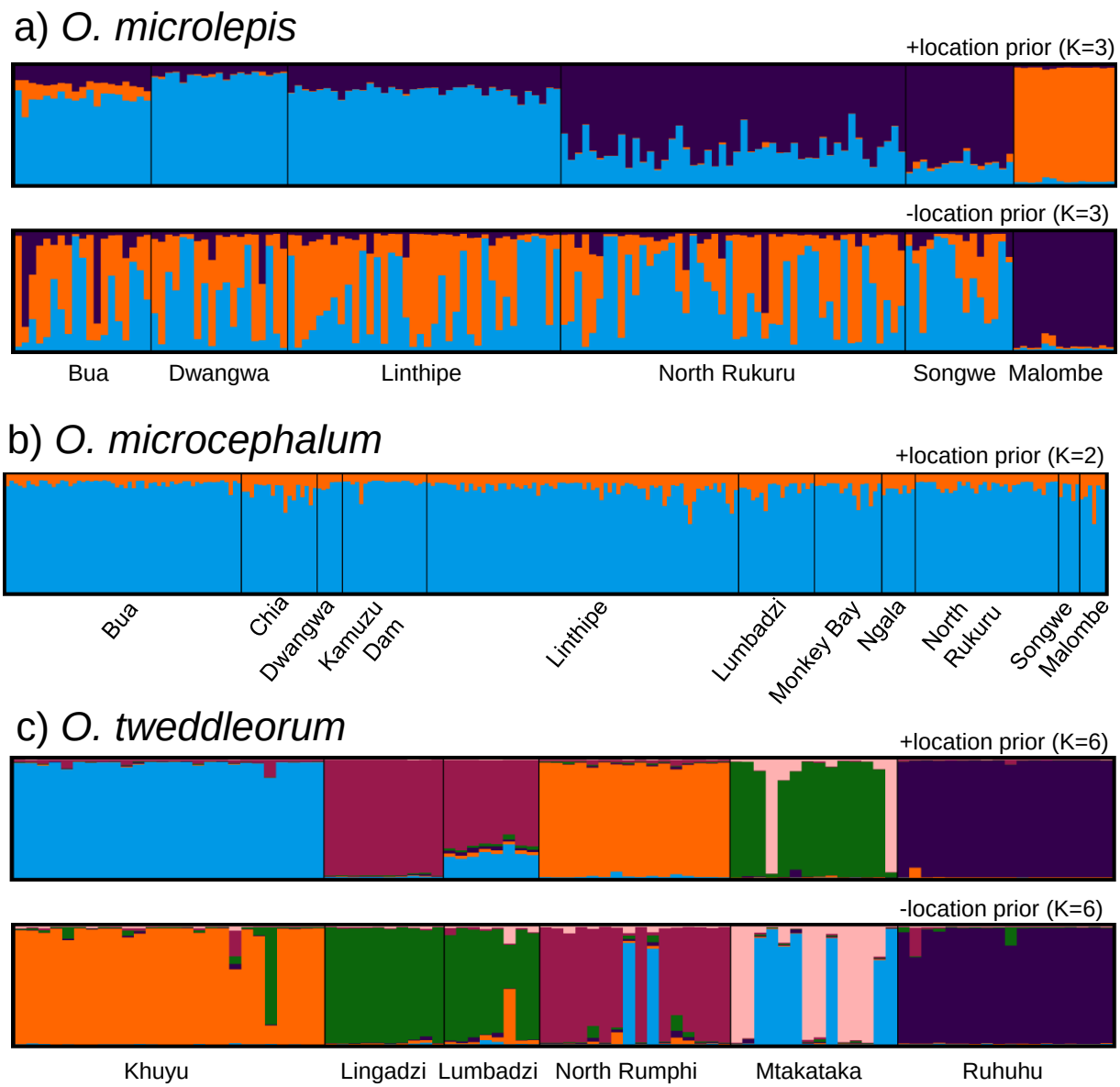

a) O. microlepis
+location prior (K=3)
-location prior (K=3)
Bua
Dwangwa
Linthipe
North Rukuru
Songwe
Malombe
b) O. microcephalum
+location prior (K=2)
Bua
Chia
Dwangwa
Kamuzu Dam
Linthipe
Lumbadzi
Monkey Bay
Ngala
North Rukuru
Songwe
Malombe
c) O. tweddleorum
+location prior (K=6)
-location prior (K=6)
Khuyu
Lingadzi
Lumbadzi
North Rumphi
Mtakataka
Ruhuhu

Fig. 3

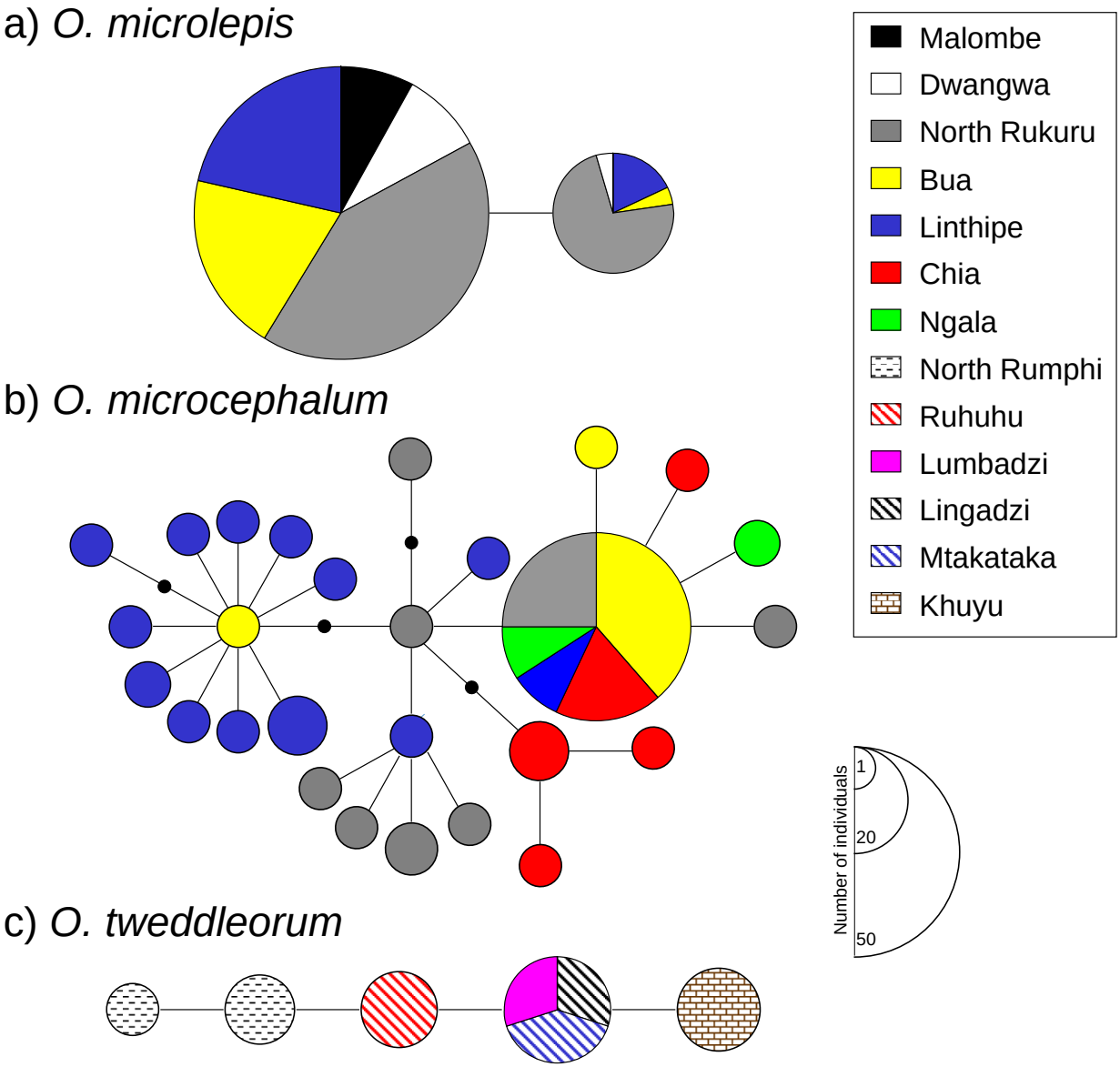

a) O. microlepis
b) O. microcephalum
c) O. tweddleorum
Malombe
Dwangwa
North Rukuru
Bua
Linthipe
Chia
Ngala
North Rumphi
Ruhuhu
Lumbadzi
Lingadzi
Mtakataka
Khuyu
Number of individuals
1
20
50

Fig. 4